DRUG SAFETY EVALUATION

Farrel L. Fort
TAP Pharmaceutical Products, Inc., Lake Forest, Illinois, U.S.A.

INTRODUCTION

Definition and Goals of Drug Safety Evaluation

Drug safety evaluation is the determination of the safety of drugs for use as therapeutic agents in humans or animals through the conduct of laboratory studies in animals and in vitro systems.

Drug safety evaluation studies are essential to

1. Differentiate between new drug entities that are unacceptably toxic and those that are not
2. Characterize the potential toxic effects of new drugs
3. Determine animal dosage levels that do not cause unwanted side effects and to estimate safe dosages to be used in clinical studies
4. Demonstrate the safety of pharmaceutical compounds or new drug formulations for marketing purposes
5. Support the marketing approval for new drugs by regulatory agencies.

Use of Drug Safety Evaluation Data

Drug safety data are used by scientists and management in the pharmaceutical industry to aid in making decisions about new drugs that can be chosen for further development. Drug safety data are necessary for approval of the Notice of Claimed Investigational Exemption for a New Drug (IND), which is required before clinical trials can begin. Clinicians conducting clinical trials refer to drug safety studies in animals for information on the side effects to be expected from the new drug and for guidance in selecting safe dosages to be used in the clinical trials. Extensive toxicology studies in animals are required for the New Drug Application (NDA) for marketing approval. Even after the marketing launch of the new drug, toxicology studies may be conducted to investigate new side effects reported after widespread clinical use. If new derivatives or new formulations of a marketed drug are discovered, toxicology studies may be conducted to determine whether the new derivative or formulation offers improved safety over the original product.

HISTORICAL PERSPECTIVE

Origins of Drug Safety Evaluation

Drug safety evaluation, as we know it today, is a relatively recent development. It probably began in the late 1930s in the United States as a result of several catastrophic events that illustrated to the public and to the Congress the need for safety testing on pharmaceutical products. These events stimulated legislation leading to regulations requiring animal safety tests before marketing of a new drug would be allowed.

Significant Events Stimulating Development of Drug Safety Evaluation

There have been several famous examples of marketed pharmaceutical products that poisoned hundreds or thousands of people (1). These examples focused the attention of the public, Congress, and the medical community on the need for safety testing of new pharmaceuticals before marketing.

During prohibition, alcoholic extracts of jamaica ginger became popular as an illicit beverage due to their 60 to 80% alcohol content. These extracts usually contained other ingredients, often including castor oil. One manufacturer in Brooklyn, New York, tried to avoid the rising price of castor oil by using triorthocresyl phosphate (TOCP) instead. TOCP is now well known to cause a delayed axonal neuropathy that results in paralysis or impaired use of the legs (known as "jamaica ginger paralysis," "ginger jake paralysis," or "jake leg paralysis"). Approximately 50,000 Americans were affected from 1930 to 1931.

In 1937, an oral formulation of sulfanilamide (sulfanilamide elixir), containing about 72% diethylene glycol, was marketed. Until that time, a suitable vehicle for an oral formulation of sulfanilamide was not available. The manufacturer chose diethylene glycol because the drug exhibited good solubility in that vehicle. Diethylene glycol is now known to cause kidney and liver damage when ingested in relatively large amounts, as occurred with sulfanilamide elixir.

About 350 people were poisoned by sulfanilamide elixir, and 105 died.

During the 1930s, dinitrophenol was touted as a weight-reduction drug. Dinitrophenol acts to uncouple oxidative phosphorylation from electron transport in the mitochondria, thereby stimulating the uncontrolled consumption of metabolic energy. This results in consumption of metabolic energy reserves, for example, fat. It also results in an elevation of body temperature. At least one person died from severe elevation of body temperature after taking an overdose of dinitrophenol. Dinitrophenol also causes cataracts; several hundred people developed cataracts as a result of taking dinitrophenol for weight reduction.

Thalidomide is the most publicized example of a disaster resulting from inadequate testing of a pharmaceutical prior to marketing. Thalidomide was never marketed in the United States but was very popular in other parts of the world, where it was taken as a sleeping aid and for colds, coughs, flu, nervousness, headaches, and asthma. It was used during pregnancy to control nausea. It was eventually reported that thalidomide causes a polyneuritis resulting in sensory and motor disturbances of the hand and thumb. However, it is best known for the induction of birth defects (e.g., phocomelia and amelia). During the late 1950s and early 1960s, over 10,000 babies were born deformed as a result of their mother's use of thalidomide during pregnancy.

Regulatory Events Influencing Development of Drug Safety Evaluation

In the United States, the regulatory environment allowing the development of drug safety evaluation as a distinct discipline began with the Food and Drug Act of 1906. This act provided governmental control over food contaminants, but provided no authority to ensure the safety of new drugs.

About the time of the sulfanilamide elixir episode, a senate bill, which was to become the new Food and Drug Act, was tied up in the House of Representatives. The publicity over the sulfanilamide elixir episode spurred Congress to pass the bill, which was signed into law in June 1938. This act created the Food and Drug Administration (FDA) and required that manufacturers of new drugs prove the product's safety to the FDA prior to marketing.

The thalidomide tragedy and the publicity it generated led to the passage in 1962 of the Kefauver–Harris Drug Amendments establishing the Investigation of New Drugs–New Drug Application (IND/NDA) system in the United States. These amendments require that before clinical studies are initiated with a new drug, the FDA has to be informed of the intent to conduct clinical studies, the nature of the clinical studies, the qualifications of the clinical investigators, and the preclinical safety studies in animals demonstrating the safety of the new drug. This information is supplied to the FDA in the form of an IND application. The 1962 amendments also require proof of efficacy before a new drug is approved for marketing. This proof is obtained by conducting clinical trials and is submitted to the FDA as part of the NDA for marketing approval (1, 2).

Influence of the International Marketplace on Drug Safety Evaluation

The international pharmaceutical market has a tremendous financial impact on all the major pharmaceutical houses. Therefore, drug safety animal studies must be conducted to satisfy the requirements of regulatory agencies around the world. In practical terms, this means satisfying the requirements of the United States, Japan, the European Economic Community (EEC), and Canada because all the other countries in the world have less stringent requirements. In certain respects, the requirements of Japan and EEC are more stringent than those of the United States. Recently, the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) has developed guidances recommending international standards for nonclinical safety studies. These ICH guidelines serve to reduce differences among regional requirements and to reduce duplication of efforts needed to register a compound internationally. However, most countries not listed above approve new drugs for marketing if those drugs have previously been approved in the United States.

The development of regulations governing drug safety studies around the world has generally followed the development of regulations in the U.S. Such regulations were promulgated in the EEC in 1980, in Canada in 1981, and in Japan in 1984.

GENERAL PRINCIPLES

There are a few general principles followed in conducting drug safety animal studies in the development of new drugs. These principles and the reasoning behind them are sometimes misunderstood by those not involved in drug

safety evaluation, leading to unwarranted criticism of toxicology studies.

First, animal toxicology studies almost always use dosage levels higher than the intended clinical dosage. One of the goals of drug safety toxicology studies is to characterize the toxic side effects of a new drug. To characterize the toxicity of a new drug, high-enough dosages must be used to allow full expression of the toxic effects of the drug. The first goal stated in the introduction was to differentiate between drugs that are acceptably safe and those that are unacceptably toxic. One criterion by which this differentiation is made is the safety margin between the expected clinical dosage, or the pharmacologically active dosage, and the dosage that causes toxicity in animals. To determine this safey margin, dosages higher than the clinical dosage must be used. In addition, dosages in laboratory animals required to elicit pharmacological and toxicological responses are usually higher than those in humans, another reason why high dosages are used in animal toxicology studies.

Margin of safety is not the only criterion used to determine if a drug is safe enough to use in humans. The type of toxicity caused by the drug and the intended clinical indication are also considered. For example, a drug that might cause cancer would not be used for symptomatic relief of a head cold. However, such a drug might be used for the treatment of cancer as the disease to be treated is itself life-threatening. Many anticancer agents fall into this category. In other words, the benefit-to-risk ratio would be too low for the head cold but high enough for the treatment of cancer. This is an extreme comparison. Another example would be the teratogenic anticonvulsants in use today. These drugs can cause malformations in fetuses. Having a seizure during pregnancy also carries serious risk to the fetus and to the mother, however. The benefits of using an anticonvulsant during pregnancy are often considered to outweight the risk of having a malformed child.

Reversibility of toxic effects is an important consideration in evaluating the safety of a drug. Cancer and fetal malformation are irreversible changes. If, instead of cancer a drug caused reversible enlargement of the liver, such a drug might be acceptable for less life-threatening diseases. In the case of the anticonvulsant used during pregnancy, a reversible retardation of neonatal growth would obviously be a more acceptable side effect than a malformation.

During the development of new drugs, reversibility of toxic effects is often studied to provide information for use in making benefit-to-risk decisions. The information is also helpful to the clinicians conducting clinical trials or using the drug if it eventually reaches the market. For the clinician it is helpful to know that if a particular side effect appears in study subjects or patients, it will disappear upon cessation of treatment. The margin of safety derived from animal toxicity studies is not always infallible. During the clinical development program, it is often necessary to increase the dosage beyond that used in initial clinical studies. In this situation, where the margin of safety is decreased because the dosage has been increased, side effects are more likely to be encountered. If those side effects have been shown to be reversible in animal studies, the clinical investigators may feel more comfortable about raising the clinical dosage in order to achieve the desired pharmacological effect.

The conduct of animal toxicity studies proceeds concurrently with and in advance of clinical studies. A certain amount of animal toxicity data is required for the IND. This often includes acute toxicity in two rodent species, mutagenicity screening studies, and one-month toxicity studies in a rodent and nonrodent species. After the clinical studies have begun, further animal studies are conducted, usually until the time for the NDA. These would include further mutagenicity studies, reproductive toxicity studies, and long-term toxicity studies, possibly including carcinogenicity studies.

The duration of treatment in animal studies is always at least as long as the duration of treatment in clinical studies and, at the later stages of development, longer than in the intended therapeutic indication. This provides another element of safety. With long-term treatment, toxicity usually occurs at lower dosages than with short-term treatment. Some drugs may be found to accumulate in the body or in certain organs and thereby give rise to toxic changes at lower dosages than with short-term treatment. Occasionally with long-term treatment, different side effects may manifest themselves. If long-term toxicity studies are conducted in advance of long-term clinical trials, any changes in safety margin or toxicity profile will be known before the conduct of those clinical trials. The duration of treatment required by regulatory agencies for marketing approval is usually longer than the intended clinical regimen.

Another principle that is generally followed in drug safety evaluation is the use of two or more species to study the toxicology of a new drug. It is usually not possible to know which species will be the most accurate predictor of the toxicological response in humans. However, if a particular side effect is found in two or more species, we can predict that it will also occur in humans at high-enough dosages. In the case where a side effect is found at a lower dosage in one species than in another, the more sensitive species is generally used to estimate the safe dosage level for humans.

TYPES OF DRUG SAFETY EVALUATION STUDIES

Acute Toxicity

Acute toxicity studies are often misunderstood and criticized for the use of large numbers of animals to generate relatively valueless data, such as LD50 (a dosage that kills 50% of the test animals). It is worth pointing out that there are legitimate uses for such data. For example, in the early stages of new drug discovery, when no toxicity data of any kind are available, it is often very helpful to determine an LD50 for comparison with the ED50 (a dose that produces a pharmacological effect in 50% of the test subjects). This comparison is often used to calculate a therapeutic index (LD50/ED50): the higher the therapeutic index, the greater the implied margin of safety between the pharmacologically active and toxic dosages. These numbers can be generated quickly and inexpensively, helping to provide an early indication of the acceptability of a new drug for further development. Another way in which LD50 data are useful is in the quality testing of new batches of a drug. In mice, LD50 data are often routinely obtained for each new batch of bulk drug for comparison with the LD50 obtained on the original batch(es) of the same drug. This quick and inexpensive test gives a final assurance that the drug is not substantially different from the original drug, or has not been contaminated prior to preparation of the final formulation. To date, there are no good substitutes for these in vivo tests. Although they do consume large numbers of animals, the tests are quick, inexpensive, and reliable.

In addition to the applications of acute toxicity studies mentioned previously, such studies are necessary as the initial toxicity studies conducted in the development of a new drug. The information provided is used to select dosages for subacute and subchronic toxicity studies in which animals are treated daily for various periods of time. Studying the acute toxicity of drug combinations is a convenient method for detecting potential drug interactions that may occur clinically. If such interactions are detected, they may be studied further in subacute and subchronic toxicity studies. In these studies, a single dose of test compound is given to animals, which are then observed for a period of 1 or 2 weeks afterward for the development of toxic signs. Surviving animals are usually necropsied at the end of the observation period to determine if grossly observable organ changes have occurred. When organ changes are observed, these are usually examined microscopically to determine the nature of the change. If an LD50 is determined, it is usually calculated using probit analysis (3). Less often, the up-and-down (4) or moving averages (5) methods, which require fewer animals, will be used. An LD50 value is not necessary for the safety evaluation of a drug or for choosing dosages for subacute toxicity studies. For these purposes, it is only necessary to determine the lowest lethal dosage of the compound and to characterize the toxic manifestations of an acute overdose. This information is useful to clinicians using the drug and to poison-control centers where information regarding the effects of acute overdose must be made available. To select dosages for repeated-dose, subacute toxicity studies, information about the lowest toxic dose is needed.

Often, an LD50 cannot be determined if the compound has low toxicity. Sometimes lethality or toxicity cannot be elicited. In these cases a "limit" dose concept is used. A limit dose, for example, 5 g/kg for rodents or 2 g/kg for dogs, is administered. If no deaths or signs of toxicity occur, the lethal or toxic dose is simply stated to be greater than the respective limit dose. Sometimes the limit dose may be determined by the volume that can be administered in a single bolus. In rodents, for example, 50 mL/kg is the maximum that can be given as a bolus orally or intravenously (IV). For IV acute toxicity, much higher dose volumes can be given if the drug solution is administered as a slow infusion.

The animals used in acute toxicity studies are usually mice, rats, and dogs. The batch tests mentioned above are usually conducted in either male or female mice. Females are generally more sensitive than are males, but either sex is appropriate. Because the objective is simply to compare the toxicity of a new batch of a given drug to that of previous batches, the absolute toxicity is less important than the toxicity relative to previous batches of the same drug. For the safety evaluation of a new drug, acute toxicity studies are usually conducted in both sexes of mice and rats by several routes of administration, including the intended clinical route(s). This is to generate initial information regarding differences in degree or type of toxicity resulting from different routes of administration [e.g., oral vs. IV vs. subcutaneous (SC) vs. intraperitoneal (IP)]. An initial indication of the degree of absorption can be obtained by comparing IV and oral toxicity. Sex differences in acute toxicity can alert the toxicologist to anticipate such differences in subacute/ subchronic studies. If sex differences are marked, this may be taken into consideration in deciding whether to pursue development of the compound. Because subacute and subchronic studies are usually conducted in rats and dogs, it is usually necessary to generate acute toxicity data in these species to aid in dosage selection for those studies.

There is increasing awareness in the industry regarding the need to reduce the numbers of animals used in acute

toxicity studies. This is because of social concerns about the unnecessary use of animals in laboratory studies and because of the financial impact of purchasing and housing large numbers of animals. The use of limit tests as outlined above and of alternative methods for calculating the LD50 offers practical means for reducing the numbers of animals used for acute toxicity studies.

Ocular and Dermal Irritation

If a drug is intended for use by ocular or dermal administration, one of the first requirements is to determine whether it could cause irritation upon contact with those tissues. Drugs or drug formulations that may inadvertently come into contact with the eye should also be tested for ocular irritation. For instance, a topical antibiotic that might be applied to the face could accidentally get into the eyes. Inhalation anesthetics also have the potential for ocular exposure.[a]

Muscle and Vein Irritation

Muscle and vein irritation studies are conducted if a drug formulation is to be administered by these routes. These studies may be conducted as a guide during the development of formulations for these routes. They should be conducted prior to any subacute studies so that experimental animals are not subjected to avoidable pain and suffering.[b]

Pain on Injection

It seems obvious that drug formulations to be given by injection should be tested for the amount of pain produced upon injection before being tested in humans. In practice, however, this kind of testing is rarely done unless a formulation is first tested in humans and found to be painful upon injection. The muscle and vein irritation studies mentioned previously sometimes give indications of pain if the animals show signs of discomfort during the injection. However, this is usually not a reliable indicator of the painfulness of the formulation. Muscle and vein irritation studies give information on the potential for drug formulations to cause tissue injury but not for pain on injection.

Models that do assess pain on injection are the rat paw lick and mouse scratch models. In the paw lick model, the drug formulation is injected into the pad of the hindfoot, and the number of times the rat licks the paw is counted (6). The mouse scratch test is similar except that the formulation is injected subcutaneously on the back, and the number of times the mouse scratches the injection site is counted.

Results from the rat paw lick model seem to correlate well with clinical painfulness experience and the muscle irritation test with certain antibiotic formulations (6, 7). However, because rat paw lick results from a wide variety of drug formulations have not been published, these correlations may hold only for a certain subset of painful formulations. A formulation could cause pain but yet not cause muscle damage upon intramuscular injection. Additionally, some formulations are painful upon intravenous injection.

Subacute, Subchronic, and Chronic Toxicity

These studies are designed to characterize the toxic effects of drugs upon repeated daily administration for periods of time ranging from 2 weeks to 1 year and to determine no-toxic-effect dosage levels for short to long-term repeated dosing.

The studies are usually conducted in rats and dogs, unless one of these species is found to be inappropriate. This may be because of a sensitivity or insensitivity to the pharmacological effects of the drug which is not expected to be characteristic of the response in humans. The rat or dog may metabolize the drug substantially differently from humans. These animals may absorb the drug poorly relative to other species, or the pharmacokinetics may be substantially different. Such differences in pharmacological responsiveness, pharmacokinetics, or metabolism are almost never known when the initial toxicology studies are conducted. An exception would be the case of certain monoclonal antibody products with binding sites specific to primate tissues. In this case, primates would be selected as the only species in which to conduct safety evaluation studies. Genetically engineered human protein products where the antigenicity to primates is expected to be less than that to other species less closely related to humans, or where the pharmacological responsiveness in primates is similar to that expected in humans, might also be tested only in nonhuman primates. In most cases, the toxicology program is conducted in rats and dogs until one of these species is found to be inappropriate. The information that might lead to this conclusion would not be available until the Phase I clinical trials are conducted in healthy volunteers.

A certain amount of toxicology data is necessary, usually including 1-month studies in rodents and

[a] For typical ocular and dermal irritation protocols, *see Animals in Drug Development*, page 000; 00000.

[b] Typical protocols for these studies are described in the article, Animals in Drug Development, in this encyclopedia.

nonrodents, prior to the initiation of Phase I clinical trials. The studies are usually conducted in a sequential fashion, using the results from the short-term studies to design subsequent studies of longer duration. Thus, acute toxicity results are used to design 2-week studies, the results of which are used to design 1-month studies, which lead to 3-month studies, and then to studies of 6 month or 1 year. A typical study design would include a period of 2 weeks or more for quarantine and acclimation of the animals to the laboratory environment. During this time, measurements can be made (e.g., body weight, food and water consumption, urine output, serum chemistries, hematology, urinalysis, electrocardiograms (ECG), and general physical and ophthalmological examinations) to determine that the animals are healthy before beginning the treatment phase of the study. Excess animals are ordered and allowed to acclimate so that after unsuitable animals are rejected, a sufficient number of healthy animals remain to begin the study.

During the quarantine period, it is customary to have the animals examined by a veterinarian for signs of disease. In some cases, particularly in preparation for a long-term study where the financial risk is great, some of the animals from the shipment may be sacrificed, blood collected for serology to determine if viral antibodies are present, and the tissues examined for evidence of infectious processes.

During the conduct of the study, all the animals are carefully monitored for changes in physical appearance, behavior, body weight, and food consumption. These variables are often the most sensitive indicators of toxicity. During the study, the other variables mentioned above are periodically measured. The variables to be measured, the intervals between measurements, and the number of such measurements during the study depend on the drug under development, the length of the study, and on the species being used. For example, when dogs are used, blood samples for serum chemistry, hematology, or drug concentration determinations can be taken more frequently because of the larger blood volume relative to rodents and the ease of obtaining blood samples. However, if interim blood samples are required in rodents, extra animals are commonly included in each group to be used for this purpose. Because of the small blood volume in a rodent and the trauma inflicted when blood is collected, there is the possibility that the process of repeated blood collection could alter the response of the test animal to the toxic insult from the drug. This possibility will vary with the frequency of blood collections, the method of blood collection (e.g., tail bleeding, retro-orbital sinus bleeding, or cardiac puncture), and the nature of the drug under study. Many investigators prefer to avoid the possibility of such confounding effects by assigning extra animals to provide blood samples. In a 1-month study, blood samples might be taken only during acclimation and at the end of the dosing period. However, if the time course of certain expected changes is of interest, then additional blood and/or urine samples may be taken after 2 weeks of treatment. In a 3-month study, samples might be taken at the end of every month or once every 6 weeks. In a 6-month study, samples might be taken every 3 months; in a 1-year study, once every 6 months. If a recovery period is included in the study design, then blood and/or urine samples would be taken at the end of the recovery period. If the recovery period is relatively long in duration, or if the time course of recovery of some function is important, samples may be taken at intervals during the recovery period. A recovery period is a period of time in which some of the animals in a study are held after the cessation of treatment to study the reversibility of any toxic changes that manifested during the treatment period. The length of recovery depends upon the nature of the toxic change under study, the duration of the treatment phase, and, to some extent, the species. For example, if the side effect caused by the drug results from an exaggerated pharmacological effect, for example, central nervous system (CNS) stimulation, cardiac arrhythmia, or hypertension, which might reverse fairly rapidly after the cessation of treatment, then a short recovery period would be appropriate. On the other hand, a change that might take longer to reverse, for example, testicular atrophy or hepatomegaly, would require a longer recovery period. In the absence of such qualifying circumstances (i.e., when the nature of the toxic changes to be expected is not known), typical recovery periods would be 2 weeks after 1 month of treatment, 1 month after 3 months or 6 months of treatment, and 3 months after 1 year of treatment.

Electrocardiograms are commonly recorded in dog studies but only rarely in rodents, although there is no good reason not to record ECGs in rodents. With the dorsal-axial-inferior (DAI) lead system (8), it is actually easier and faster to record ECGs in rats than in dogs. Furthermore, the rat heart is known to respond to pharmacological agents (except the cardiac glycosides) affecting the ECG of other species (8). Because of the more rapid heart rate of rat, the ECG is slightly more difficult to evaluate. However, the main reasons why ECGs are not recorded more commonly in rats in toxicology studies are custom and historical precedent.

It is also fairly easy to perform specialized measurements such as electroretinography in dogs without the anesthesia or surgical preparation required in other species (9, 10). Electroretinography has assumed a status of increased importance in safety evaluation because many

drugs affect retinal function. Examples are the quinolone antibacterials and Viagra (11).

At the end of the study, all the animals are killed under anesthesia, the internal organs are examined for visible changes indicative of pathological responses, the organs are weighed, and samples of all tissues are taken for histological examination. In some studies, particularly long-term studies, animals from each group may be killed at intervals during the study to collect tissues for histological examination. For example, in a 6-month study, some animals may be killed after 3 months of treatment; in a 1-year study, some animals may be killed after 6 months of treatment. Animals would also be killed at the end of any recovery period, or at intervals during the recovery period, for assessment of drug-related tissue changes. In studies with large animals (e.g., dogs or monkeys), protocols are sometimes written so that all the tissues from all the animals are examined histologically. However, in rodent toxicity studies, because of the large number of animals used, protocols are commonly written so that tissues from animals in the control and highest dosage groups killed at the end of the treatment period are examined first. If any changes are found that appear to be drug related, tissues from animals in the next lower dosage group are examined, and so on. Tissues from recovery group animals are usually examined only in groups where drug-related changes are found at the end of treatment.

The determination of the no-toxic-effect dosage level in a toxicity study is not always easy; often the only findings at a particular dosage level are those resulting from the pharmacological action of the drug, for example, hypotension and the resultant behavioral signs with a peripheral vasoditator or sedation with a tranquilizer. Are these changes to be considered due to toxicity? Certainly, they would be unwanted side effects of the drug, but on the other hand, they would not be unexpected based on the pharmacology of the drug. Other examples would be hepatocellular hypertrophy with a drug that also causes induction of liver microsomal enzymes, or renal enlargement at high dosages. These changes might be considered adaptive physiological changes. Should they be considered to be due to toxicity? Some investigators look for a slightly different end point in determining the results of a study—the no-observed-effect dosage level. In their view, all the examples just cited would be lumped along with other findings more clearly labeled toxicity into the category of drug-induced observed effects. They look for a dosage level below which no drug-induced effects are observed.

Dosage selection for subacute, subchronic, and chronic studies is always a sensitive matter, requiring experience and judgment. Three dosage levels plus one or more vehicle or untreated control groups are usually used. The highest dosage is selected to elicit some degree of toxicity. It should not be so toxic that animals die or are debilitated before the end of the study. The purpose is to characterize the toxicity of the drug on repeated administration. If the dosage is too high, animals may be lost before the end of the treatment period. This reduces the number of animals available for terminal serum chemistry, hematology, histology, and so on, so that the toxicity cannot be characterized adequately. On the other hand, if the dosage is not high enough to cause toxicity, the study may be rejected as inadequate by regulatory agencies. The low dosage is usually selected to provide some small multiple of the expected clinical dosage or systemic plasma exposure level. In cases where the drug is fairly toxic, the low dosage may be slected to match the clinical dosage or plasma exposure level. However, it is usually considered necessary to demonstrate a no-toxic-effect dosage of at least 10 or 6 times the clinical dosage in the case of rats or dogs, respectively. Therefore, the low dosage is usually at least this multiple of the clinical dosage. The mid-dosage is selected to be at equal log intervals between the high and low dosages. This is because the response of biological systems to the vast majority of stimuli are logarithmic. Therefore, to best demonstrate a linear dose–response relationship, the stimulus should be administered on a logarithmic scale. In selecting the dosages for a study, it must be remembered that if the high dosage is toxic as expected, the mid-dosage is likely to be the no-toxic-effect dosage. Therefore, the relationship of the mid-dosage to the clinical dosage should also be considered when selecting the dosage scheme for a study. Often, it is desirable to compare the toxicity in long-term studies with the same dosages in short-term studies. If this is the case, some of the dosage levels in long-term studies may be the same as those in short-term studies. Although three dosage levels are usually considered the minimum, sometimes more dosage levels are used to meet all the considerations discussed above.

The number of animals per group depends on the length of the study, the species, whether a recovery period will be included, and whether extra animals will be included for blood sampling or other special procedures. Generally, for rodent studies to be submitted to a regulatory agency, 10 per sex is the minimum for studies up to 3 months, and 20 per sex is the minimum for longer studies up to 1 year. For dogs, at least 4 per sex is usually required in studies submitted for registration. More animals are preferred for long-term studies because the longer the study duration, the greater the chance of losing animals due to accidents or disease. The study design should compensate for the greater statistical chance of losing animals so that an

adequate number of animals are available at the end of the study for evaluation.

The route of administration is an important consideration in the design of a study. Generally, the route of administration in drug safety evaluation studies is the same route intended clinically. The most common route is oral. Drugs can be administered easily by gavage in rats, dogs, and monkeys and by capsule in dogs. Methods are available for administering capsules to monkeys and rats. In these cases, the capsule is usually affixed to the end of a special gavage needle for rats or gavage tube for monkeys. The tube is inserted into the stomach and the capsule expelled from the end of the tube. In long term (i.e., 1 year) rodent studies, it is common to administer drug mixed in the diet. This assumes that the drug is stable and bioavailable when mixed in the diet and that the drug can be mixed in a homogeneous manner in the diet. All these requirements must be demonstrated before committing to a dietary administration study. When administering drugs via the diet, it is necessary to adjust the drug concentrations in the diet to compensate for variations in food consumption by the animals and for changes in body weight. During the early stages, for example, the first 3 months, of the study when the animals grow most rapidly and when drug effects on food consumption and body weight gain are likely to be most pronounced, this should be done on a weekly basis. Later in the study, say, after 3 months, when the growth curve of rodents approaches a shallow slope, this can be done once per month. This is done by adjusting the concentration of drug in diet based on the food consumption and body weight gain from the previous period. The concentration for each sex group is adjusted separately. If an intravenous study is needed, this could limit the feasible duration of the study and/or the number of animals or groups that can be accomodated. If the drug formulation does not cause vein irritation, it is feasible to conduct 3-month studies in rats, using the lateral tail veins for injection. In dogs, almost any length of study should be possible; however, 3- or 6-month studies seem to be within the usual limits of practicability. If the drug must be given by slow injection rather than by bolus or slow push, there will be limitations on the numbers of groups and animals that can be used due to the equipment and increased technical staff needed to conduct such studies.

Mutagenicity

Mutagenicity studies are required for the marketing approval of new drugs. The role of mutagenicity studies in drug safety evaluation is sometimes misunderstood. The purpose of these studies is not to replace carcinogenicity bioassays but to predict which drugs are likely to be positive in carcinogenicity bioassays and cause cancer in humans and to detect compounds that might cause hereditable germ-line mutations in humans. Such predictions require an in-depth knowledge of the mutagenicity assays available and scientific judgment. Sound predictions cannot be made strictly on the basis of numbers of positive or negative results in a few tests.

It is currently accepted knowledge that many, but not all, known carcinogens have the ability to cause damage to deoxyribonucleic acid (DNA). Such agents also have the potential to cause heritable germ line mutations. The mutagenicity studies currently available were designed to detect DNA-damaging agents. Positive results in these tests therefore indicate the possibility of carcinogenicity and germ cell mutagenicity resulting from the test compound. However, a negative result in mutagenicity tests designed to detect DNA-damaging agents does not imply that the compound will not be a carcinogen. We now realize that some carcinogens are nongenotoxic. Examples are the rodent liver carcinogens, phenobarbital and chlorophenothane (DDT), which are not mutagenic. These compounds are thought to cause liver tumors through their action as promoters. Bile acids are known to act as promoters in the induction of intestinal tumors. In experimental situations, certain compounds are known to act as cocarcinogens (enhancing the carcinogenic response only when applied concurrently with a DNA-damaging agent) or as promoters (enhancing the carcinogenic response when applied after the DNA-damaging agent). These types of carcinogens are not detected by mutagenicity assays. Currently, there are no short-term assays available for routine testing which reliably detect cocarcinogens or promoters, that is, nongenotoxic carcinogens.

Much publicity has been given to the absence of a high level of concordance among the results of a selected battery of four in vitro mutagenicity assays with the results of carcinogenicity bioassays in rats and mice (12). First, the number of compounds available for comparison is hardly an adequate sample size from which to make any general conclusions regarding the predictive capabilities of any series of assays. Second, the compounds selected for comparison of bioassays to genetic toxicity assays were heavily weighted in favor of genotoxic compounds. This makes the estimation of specificity for the mutagenicity assays (the ability to produce a negative result with a true noncarcinogen) difficult. Both sensitivity (the ability to produce a positive result with a true carcinogen) and specificity are used to determine concordance (the ability to produce a correct response with either a true carcinogen or a true noncarcinogen).

Furthermore, the compounds selected surely included compounds that were carcinogenic in the rodent bioassays but are not genotoxic. Finally, the four mutagenicity assays selected represent an artificial restriction on the comparison of mutagenicity assays to bioassay results. In practice, the genetic toxicologist is not restricted to using only a few specific assays.

As stated at the beginning of this section, certain mutagenicity tests are required by regulatory agencies for approval of new drugs. The ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use) (13) requires a battery of three categories of tests: (1) bacterial mutagenicity, (2) in vitro cytogenetics or in vitro mammalian cell mutagenicity assay with ability to detect chromosomal damage (i.e., mouse lymphoma tk [thymidine kinase locus] assay), and (3) in vivo test for chromosomal damage using rodent hematopoietic cells. For initial screening of new compounds, I recommend bacterial mutagenicity and in vitro cytogenetics. This is because the bacterial mutagenicity test, usually the Ames test using *Salmonella typhimurium*, is rapid, inexpensive, and sensitive. In vitro cytogenetics should be included because the bacterial test cannot detect agents that act at the level of the eukaryotic chromosome. Therefore, using these two tests, one can determine relatively quickly whether a new drug is a potent mutagen or not. If positive results are obtained in in vitro studies, other tests can be conducted to try to determine the mechanism responsible for the positive results, and to determine the potency and any specificity of the mutagenic response. If the drug is positive for mutagenicity in some tests but has relatively low potency, the drug might still be developed. In this case, in vivo mutagenicity tests may reveal a lack of detectable response. This would aid in defending the benefit/risk ratio for the drug. Some drugs may be positive in in vitro mutagenicity tests only in the presence of activating microsomal preparations. If this is the case, analysis of the metabolic products of the in vitro reaction may reveal that those products are not the same as the metabolites produced in vivo. Thus, the relevance of the in vitro mutagenicity study to in vivo risk assessment can be determined. The point is that one should not just accept in vitro study results at face value. Further study of positive results is usually warranted. Also, some of the mammalian cell in vitro mutagenicity assays are known to be sensitive to alterations in the osmolarity and/or the pH of the culture medium. Examples are the in vitro cytogenetics in Chinese hamster ovary cells and the L5178Y mouse lymphoma assays. The osmolarity in the assays should not be greater than 600 milliosmolar (14) and the pH should be in the range of 6.8–8.0 (15).

Several transgenic animal models are available for mutagenicity testing *in vivo*. The most widely used are the Big Blue™ and Muta™ Mouse models. These animals have portions of the *E. coli* galactosidase operon inserted into their genome. This bacterial DNA is used as the target for mutagenicity testing when the animals are treated with test chemicals. After chemical treatment, the bacterial DNA is removed from the tissues and packaged into bacterial viral (phage) particles. The phages are used to infect bacteria growing on agar plates. The bacteria multiply to form colonies that can be counted. Because of a color reaction that can be catalyzed by the galactosidase protein, a color reaction formed in the bacterial colonies can be used to assess the presence or absence of mutations formed in the target DNA from virtually any tissue within the treated mice.

Guidelines and protocols for specific mutagenicity assays can be found in Naismith, 1987; Brusick, 1980; Kilby et al., 1984; Mirsalis et al., 1994; Provost et al., 1993; and Heddle et al., 2000, listed in the Bibliography.

Reproductive Toxicity

Reproductive toxicology covers the entire process of reproduction from mating to pregnancy, birth, weaning, and, sometimes, the reproductive function of subsequent generations. The species most commonly used in these studies are rats and rabbits, although in special circumstances, other species such as mice, dogs, or monkeys are used. A series of studies is usually conducted so as to cover all the phases reproduction. Within each study, groups of animals corresponding to untreated controls and 2 to 3 drug-treated groups are used.

To assess potential effects on all phases of reproduction, separate studies are usually conducted that focus on distinct aspects of reproduction. The separate studies are usually designed to provide some overlap in the phases tested so as to avoid in gaps in the testing program. Fertility studies (historically called Segment-I studies) are generally intended to assess effects of the drug on the processes of spermatogenesis, oogenesis, mating behavior, copulation, fertilization, early embryonic survival and development, and implantation of the zygote in the uterus. Teratology (Segment II) studies are intended to assess effects of the drug on embryonic/fetal survival and development. Peri-/post-natal (Segment III) studies assess drug effects on late pregnancy, parturition, lactation, and postnatal development. One reason for dividing the reproductive process into phases is that, generally speaking, the shorter the treatment period, the higher the dosage that can be given without causing excessive toxicity. By dividing the reproductive process into shorter

phases instead of using the long phase required to encompass the whole process, higher dosages can be given to maximize characterization of the drug's toxicity in any part of the reproductive process. If a drug effect is found, the stage of the reproductive process affected by the drug can be more readily identified. This facilitates the assessment of risk for clinical use and helps the toxicologist design further studies, if needed, to examine the mechanism of action. A practical reason is that smaller studies can be conducted, sometimes in parallel fashion. Smaller studies allow flexibility in the timing of the evaluation of various parts of the reproductive process in accordance with the development program of a particular drug.

Fertility studies

Segment-I fertility studies are usually conducted in rats, although mice may occasionally be used. Rodents are used because of the convenience of housing large numbers of them, ease of mating, and widespread experience among reproductive toxicologists with these species. In these studies, the drug is usually administered to males for 28–30 days and to females for 14 days prior to mating. The duration of the spermatogenic cycle in rodents is approximately 60 to 80 days. However, a 28- to 30-day exposure period is considered sufficient to detect changes to the male reproduction when used in conjunction with thorough histological evaluation of the male reproductive tract. The treatment duration in females is considered sufficient to allow enough time for the females to be exposed to steady-state blood levels of the drug for at least a week to 10 days before mating. The animals are then placed with individual members of the opposite sex for mating. Treated males may be mated with treated females, or treated members of one sex may be placed with untreated members of the opposite sex. The females are observed for signs of mating, copulatory plug or sperm-positive vaginal smear, to determine mating performance. Generally, mating treated males to treated females is less costly because of the smaller numbers of animals involved. However, if drug-related effects are found, it may be necessary to repeat the study by mating treated animals with untreated animals. Drug treatment of the females is continued through the mating period and through mid-gestation. Treatment of the males continues until it is possible to determine whether there were any effects on fertility. If such effects are found, it may be desirable to remate the males to see if effects on fertility are reversible, or subject the males to histological evaluation of the reproductive tract to try to determine the cause of any fertility effects. Waiting until the end of pregnancy for the intrauterine examination offers the advantage of an additional assessment for effects on intrauterine development. The fertility study may or may not be initiated prior to the IND application; however, it is usually required for the NDA. This fertility study in male animals is conducted prior to large-scale (phase III) clinical trials in men. This study is generally conducted prior to any clinical trials involving women of child-bearing potential. However, in the U.S. and Europe, women of child-bearing potential confirmed to be non-pregnant and using effective birth control may be enrolled in phase I and phase II (usually small-scale) studies prior to completion of the female fertility study.

Teratology studies

In teratology (hisorically called segment-II) studies, a drug is administered to pregnant animals during the period of fetal organogenesis. The studies are usually conducted in rats and rabbits, although other species such as mice or monkeys, are also commonly used. Most regulatory guidelines require that teratology studies be conducted on two species, and most suggest rats and rabbits as the preferred species. These species are used because of the ease of housing and handling, the ease of mating or insemination (rabbits), their short gestation period, and their large litter sizes (which facilitates statistical evaluation of the data). The period of drug treatment during gestation can be confusing. Most guidelines recommend gestation Days 6 to 15 for mice and rats and Days 6 to 18 for rabbits. Japanese guidelines recommend Days 7 to 17 for rats. These guidelines assume that the first day of pregnancy (the day when a vaginal plug or sperm in the vagina is found) is counted as gestational day 0. However, the guidelines of the Organization of Economic Cooperation and Development (OECD) recommend that if animals are artificially inseminated or pregnancy is determined by forced mating (rabbits are usually impregnated by one of these methods), the days of treatment be calculated by adding 1 day to the usual number. To accommodate all these guidelines, I recommend administering drugs during gestation Days 6 to 17 for rodents and Days 6 to 19 for rabbits. The extra day or two of dosing will have no adverse affect on the study but will allow compliance with worldwide guidelines.

At the end of the gestation, the dams are killed and the uteri examined internally for the presence of live, dead, resorbed, or degenerating fetuses. At this time the number of implantation sites and the number of corpora lutea are counted, and the internal organs of the dam are examined visually to detect any pathological changes that may have affected the pregnancy. The pups are weighed and examined for any external abnormalities. A number of the pups are fixed and prepared for examination of the

internal organs, and the rest are fixed and prepared for skeletal examination. In the case of rabbits, all the pups are subjected to both internal and skeletal examination. The abnormalities found are classified as either malformations or developmental variations. Malformations are those changes that are irreversible, alter general body conformity, interfere with body function or are possibly life-threatening, and are not commonly found in untreated animals. Variations are those changes that may be reversible, have no effect on body conformity or the health of the animal, represent slight deviations from the normal, and are found occasionally in untreated animals. Some changes, for example, reduced skeletal ossification or enlarged renal pelvis (apparent hydronephrosis), may simply be indicative of delayed intrauterine growth.

Dosages are selected so that some maternal toxicity is demonstrated during the study. This is to ensure a thorough test of teratogenic potential. It also has a practical application when the drug is used clinically. If during drug treatment a pregnant woman develops a toxic response, the physician is alerted to reduce the dosage or withdraw treatment. If it is known that the drug does not cause fetal toxicity or malformations at dosages up to those causing maternal toxicity, the physician can be assured that the fetus remains unharmed. If, on the other hand, a drug caused toxicity or developmental abnormalities in fetuses at dosages below those causing maternal toxicity, the physician no longer has the warning sign of maternal toxicity. With such a drug, the fetus may be harmed in the absence of maternal toxicity. Obviously, this is a more dangerous situation. Thus, the concept of toxicity to the adult mother versus the developing conceptus (or A/D ratio) assumes practical importance. Often, the results of reproductive toxicology studies are reported in terms of no-toxic-effect dosage for the dam and no-toxic-effect dosage for the fetus or offspring.

The special case of teratology studies in rabbits is worth mentioning. Rabbits are very susceptible to changes in the intestinal flora caused by some antibiotics. This usually leads to a marked reduction in food consumption, which leads secondarily to fetal toxicity, abortions, and fetal malformations. In such cases, the rabbit exhibits a species-specific sensitivity to the drug and may be an inappropriate species for a teratology study. However, the standard protocol may be modified so that the dams are treated with a drug for only a few days during the gestation period (e.g., Days 6 to 13 and Days 14 to 19, or smaller intervals). In this manner, excessive toxicity to the dam can be avoided while sufficiently high dosages of the drug can be given to provide an adequate test for teratogenicity.

The teratology study is often the first study of reproductive toxicity because it is less time- and resource consuming than are studies involving other phases of reproduction and because it assesses an end point, teratogenicity, which can have a major impact on further development of a drug.

Perinatal and postnatal toxicity

The perinatal and postnatal study (sometimes called a Segment III study) is usually conducted in rats. In this study, the drug is administered to pregnant rats from implantation (gestation day 6) through parturition and lactation. If desired, some of the fetuses can be taken at the end of gestation for morphological exams to satisfy the requirements for testing effects on organogenesis (teratology or Segment II phase). Effects of the drug on fetal development, parturition, nursing, growth, development, learning/memory, emotionality, and often, reproductive function of the pups are assessed. Postnatal developmental indices such as body weight gain, time of eye opening, pinna detachment, incisor eruption, vaginal opening, and development of righting and startle reflexes are evaluated. Some of the pups are retained after the lactation period for determination of behavioral variables such as learning, memory, and emotionality. Tests such as active or passive avoidance, shuttle box, and water T or M maze assess learning and memory. Test such as activity measurements or open-field behavior give some assessment of emotionality. In addition some of the pups may be retained for mating to determine the fertility and reproductive performance of the F1 generation. In this case, the study is usually terminated at the end of the F1 generation's pregnancy, but may be extended to assess early postnatal survival of the F2 generation. Again, the rationale is to conduct a stringent test for effects of the drug on development of the offspring.

Carcinogenicity Bioassays

Carcinogenicity bioassays are long-term studies conducted in rodents to assess the tumor-producing potential of a compound. These studies are conducted according to guidelines established originally by the U.S. National Cancer Institute. However, other guidelines, for example, OECD and FDA, are also useful in developing protocols. It is required that bioassays be conducted on both rats and mice over a period of time representing a majority of the natural life span for the species. Depending on the strain used, this is usually 2 years for rats and 18 months to 2 years for mice. The animals must be no older than 6 weeks at the initiation of treatment. The strains used are

most commonly Fisher 344 or Sprague Dawley rats and B6C3F1 or CD-1 mice. The animals are dosed daily, usually with orally administered drugs. If the drug is stable in the rodent's diet and can be mixed with it homogeneously, it is commonly administered via the diet because of the reduced staff required. Again, in the case of drugs, the concentrations of test compound in the diet are usually adjusted at regular intervals in order to maintain the target dosage level. The animals are examined frequently for the presence of palpable masses. Blood samples may be taken at 6-month or 1-year intervals for hematological evaluation, with particular attention paid to the development of leukemias. At the end of the study, surviving animals are killed and their tissues subjected to a thorough evaluation for the presence of neoplasia. The study design usually includes three drug-treated groups plus one or more control groups. The number of animals is usually 50 per sex for the drug-treated groups and 100 per sex for the control groups. However, extra animals (e.g., five per sex-group) are often started on treatment for the first 6 to 8 weeks of the study so that if animals die during the early phases of the study, which commonly occurs, there will still be enough animals on study to maintain the intended group sizes. The dosages are selected on the basis of the results of 3-month maximum tolerated dosage (MTD) studies. The object of the MTD studies is to determine the MTD for each sex of each mouse and rat strain to be used in the bioassay. The MTD is defined as a dosage that causes some degree of toxicity but not a decrement in body weight gain greater than 10% compared with that of controls; also, that does not cause toxicity, other than that related to a neoplastic response, which would be expected to shorten the animal's natural life span.

Recently, transgenic strains of mice have become available, which allow detection of a carcinogenic response within 6 months of treatment rather than 2 years. Examples of these are the P53 +/− (P53 knockout), TG.AC and the rasH2 mouse models. These animals contain deletions of genes involved in suppressing the formation of tumors. The use of these animal models is justified based upon the fact that these same genes are known to be involved in many human tumors. Thus, the mechanisms leading to enhanced tumorigenic response in these animals is known to be relevant to human tumorigenesis.

There has been controversy over the use of a maximum tolerated dosage in conducting carcinogenicity bioassays. The argument against the MTD concept is that the dosages usually required to satisfy the MTD criteria are unrealistically high relative to the exposure levels expected for the drug. Thus, a carcinogenic response at the MTD would not be predictive of the same response in humans. However, it is my opinion that this high dosage is necessary to give a stringent test of the drug's potential to cause cancer. After all, we are trying to predict the response of a heterogeneous population with different susceptibilities to potential cancerogenic agents with relatively small numbers of homogeneous strains of animals. Even though we treat the animals for most of their lives, this is still only a 2-year period. Humans are potentially exposed to some drugs for much longer periods of time. We must, therefore, push the test system in order to understand the potential of the new drug. Nevertheless, there are cases where valid exceptions can be made. For instance, if the metabolism of the drug changes (different metabolites or kinetics are observed) at a particular dosage level, then the capacity of the organism to handle the drug in the usual manner has probably been exceeded. A carcinogenic response, or the lack of one, at dosages higher than that level would not be predictive. Obviously, if a drug is nontoxic, then a MTD may not be possible. If a drug causes alterations in levels of, or target organ responsiveness to, sex or thyroid hormones, high dosages may lead to tumor formation in animals secondary to the endocrine disturbances. If such disturbances do not occur below a certain threshold level, and if they are not expected at the human therapeutic dosage, tumors formed in a bioassay at higher than threshold levels will not be predictive of the risk to humans. The same argument applies if the drug stimulates cell division in the organs such as the liver of experimental animals. Such stimulation of cell division will likely lead to tumor formation in rodents. Again, dosages below those that stimulate cell division must be used for extrapolation of bioassay results to humans (16, 17).

The ICH guidelines allow several alternative ways of selecting the highest dosage for carcinogenicity studies in addition to an MTD. The highest dosage may also be selected based on dose-limiting pharmacodynamic effects, or systemic plasma exposure levels at least 25-fold higher than the human plasma. Certain pharmacodynamic effects if allowed to become exaggerated may affect the viability of animals on a study. Examples would be excessive hypotension with an agent that would be used to treat hypertension. In this case, excessive pharmacological action may lead to the death or poor condition of animals before the potential for carcinogenicity could be assessed. Other agents may cause behavioral changes. Dopamine agonists, for example, cause hyperactivity and stereotypical behaviors such as self-mutilating behaviors which, if allowed to become excessive, represent an issue in terms of humane, ethical treatment of animals. The 25-fold AUC criterion can be used only when there is no evidence for

genotoxicity. Under this condition, a dosage which gives a plasma systemic exposure level in terms of AUC (Area Under the Curve) that is 25 times higher than the maximal anticipated human systemic exposure is allowed even if toxicity is not produced. In this case AUCs of both parent drug and metabolites are considered, both individually and combined. In addition, the plasma exposures must be normalized for any differences in plasma protein binding between animals and humans. The "limit" dose concept can be used under certain conditions. There must be no evidence of genotoxicity, the human dose must be $\leq$500 mg/day, and the AUCs obtained in animals must be $\geq$10 times the human AUCs. If these conditions are met, the high dosage in a carcinogenicity study may be limited to 1500 mg/kg/day. Saturation of absorption means that at some point, if a higher dosage is given, no further increase in systemic exposure (plasma AUC) will be obtained. Under these conditions a dosage that is slightly above the dosage where the AUCs stop increasing with dosage will be accepted as the highest dosage. The maximum feasible dosage is limited by the amount of formulation that can be delivered safely to the animals without trauma (usually 10–20 ml/kg/day by oral gavage) or intefering with nutrition (e.g., 5% mixed in the diet). Other factors such as ability to prepare a homogeneous mixture with vehicle or diet or other physical characteristics of the compound or formulation may also be limiting factors.

The statistical analysis of bioassay results is sometimes controversial. Problems arise when early deaths occur and the distribution of such early deaths is not uniform among study groups. This is because in groups where more animals die early, the number of animals remaining at risk for the development of tumors over the remainder of the study is reduced. Thus, the probability of finding a positive result is artificially reduced for those groups. To overcome this problem, a method is commonly used whereby the entire study period is divided into small time intervals, and animals having tumors discovered at necropsy or tumors considered to have caused the death of the animal are compared among groups within the small time intervals. A comparison of the summation of tumor incidence rates within the time intervals is then used to determine if treated groups exhibit an increased incidence of tumors (18–20).

It is important to understand the situations and criteria that determine the need for a carcinogenicity bioassay of a new drug. If a drug is intended for long-term use, for example, some antihypertensive agents or benzodiazepine tranquilizers where the duration of exposure could be years or the remainder of one's lifetime, or if it will be available for widespread, uncontrolled use, then bioassays will likely be required. If the drug affects rapidly growing tissues, for example, bone marrow or intestinal mucosa, or is known to affect cell division, bioassays may be required. If on the other hand, the drug is intended for short-term, controlled use, say, a few days, weeks, or months, and does not affect rapidly growing tissues, then bioassays will not be required unless mutagenicity tests indicate the drug represents a risk. If positive results have been obtained in one or more mutagenicity tests, the probability is high that a bioassay will be required. If positive mutagenicity results were obtained in in vitro studies only, and appropriate in vivo studies are negative, the probability of having to conduct a bioassay is much lower than if the in vivo studies are positive. In any case, I recommend holding discussions with the appropriate regulatory authority to determine the case for the particular drug under study. Currently, the cost for a set of bioassay studies in rats and mice is close to two million dollars. Therefore, the cost involved warrants a thorough mutagenicity program and meaningful interaction with regulatory authorities.

Immunological Sensitization

New drugs are sometimes tested for antigenic sensitization if they are suspected of being sensitizers. Such drugs would have a chemical structure similar to that of known sensitizers or would be high-molecular-weight compounds such as peptides. Agents to be administered topically are generally tested for dermal sensitization regardless of chemical structure.

Typical protocols for dermal sensitization, active systemic anaphylaxis, and passive cutaneous anaphylaxis tests are given in the article Animals in Drug Development, in this encyclopedia. In a typical gel-diffusion test, solutions of test compound, egg albumin (positive control), or vehicle (usually saline) are mixed with complete Freund's adjuvant. These mixtures are injected into animals (usually guinea pigs) of the respective treatment group once a week for 3 weeks. Two weeks after the last injection, blood is collected by cardiac puncture and serum (antiserum) is prepared. The gel diffusion is carried out on double immunodiffusion (Ouchterlony) discs containing trypan blue. Each disc contains a pattern of wells consisting of a circle of six 5-mm wells around a single 6-mm well. Antiserum is placed in the center well, and test compound or positive control or vehicle is placed in the wells surrounding the center well. The discs are incubated in a humidified 37°C incubator for up to a week to allow development of precipitation bands resulting from interaction between antigens (test compound or positive control) and antibodies (in antiserum). The precipitation bands are visible to the naked eye or with a dissecting microscope (21, 22).

Phototoxicity

Phototoxicity testing is not often included in the safety assessment of new pharmaceuticals; however, such studies may be conducted if phototoxicity is expected on the basis of reports on chemicals of similar structure or pharmacological class. Information on UV absorption spectrum of a compound may prompt phototoxicity testing. Sometimes reports of phototoxicity after a drug has reached the marketplace prompt laboratory studies of phototoxicity to determine if the drug is causing phototoxicity and, if so, by what mechanism.

Phototoxicity is to be distinguished from photosensitization (or photoallergy). Phototoxicity results from interaction of the drug or drug metabolites in the skin with light, usually in the UV region, to produce reactive molecular species that cause cell injury or death in the skin. Phototoxicity is most commonly produced under sunlight where the skin is exposed to the full range of UV light. The phototoxic response generally shows a dose response to both the amount of drug and the intensity or duration of light exposure. Photosensitization, on the other hand, results from an immunological response to by-products formed in the skin when the drug or its metabolites interact with light. The initial exposure to the drug and light may not result in an adverse reaction, but subsequent exposures to drug and light will result in an allergic reaction in sensitized individuals. The models described in this section are confined to the assessment of phototoxicity. Modifications of the models for dermal sensitization referred to in the section on immunological sensitization, incorporating light (i.e., UV light) as well as drug exposure, can be used to test for photosensitization.

Several sensitive and convenient models can be used to test for phototoxicity (23, 24). Guinea pigs, rabbits, and mice are the most commonly used species. Normal, albino guinea pigs and rabbits must be shaved or their hair removed with a depilatory before light exposure. Hairless mice have been used because of the obvious advantage of not having to remove the hair prior to testing. The euthymic, hairless guinea pig has also been used. These animals provide a convenient and sensitive model for phototoxicity testing (25). When conducting a phototoxicity test with normal animals, one finds that the hair can never be completely removed from the test site, and that this interferes to varying degrees with light exposure and scoring the resulting skin reaction. In addition, the ears, a site containing relatively small skin area, are sensitive indicators of phototoxicity. In some cases, the ears may be the only part of the animal where a skin reaction is obtained. With hairless animals, a much larger skin area is available for accurate observation of the skin reaction. The larger size of the hairless guinea pig compared with that of hairless mice maximizes this advantage. With hairless animals, accurate scoring of skin reactions is facilitated because inevitably there will be an area where exposed and unexposed skin are juxtaposed along the sides of the animal. (This assumes the animal was exposed to light coming from above with the animal in the normal resting position on all four feet.) The comparison of the appearance of exposed and unexposed skin side by side in the same animal makes it easier to discern slight changes in skin appearance due to light exposure.

A test using the mouse tail has also been recommended for determining phototoxicity (26). In this test, mice of the same strain, age, sex, and weight are used. Groups of mice are exposed to the drug and UV light; after a period of 24–48 hours, depending on the drug, the tails are removed and the wet weight of the tails is determined. An increase in tail wet weight indicates an inflammatory response due to phototoxicity. Obviously, an untreated group as well as groups treated with only the drug or the light are included as controls.

In Vitro Toxicity

In vitro toxicity studies are being used more commonly in drug safety evaluation as improved techniques are being developed. Their main uses at present are as a tool for screening compounds for toxicity to particular cell types and in studying cellular mechanisms for target organ toxicity.

The most common assay systems consist of cell cultures that are treated with the drug in the culture medium. Toxicity is commonly assessed by release of intracellular enzyme, for example, lactate dehydrogenase or aspartate transaminase, into the culture medium or by other indicators such as decreases in the rate of radiolabeled amino acid or nucteotide precursor incorporation into macromolecules. A decrease in the intracellular uptake of the vital dye, neutral red, has also been used as a measure of toxicity with a variety of cell types. Fluorescent dyes are also available which allow, for example, estimations of intracellular Ca^{2+} concentrations, intracellular pH, or mitochondrial function. Such measurements can be helpful indicators of cellular toxicity and can help determine mechanism of toxicity.

An example of screening compounds for toxicity to particular cell types is the use of rat hepatocyte primary cultures in screening macrolide antibiotics for potential liver toxicity. The liver is a target organ for this class of compounds. In the development of new macrolides, this in vitro culture system is useful in discriminating those

macrolides that are more toxic to liver cells and those that are more toxic than other well-known and widely used antibiotics such as erythromycin. This information, together with data on antimicrobial potency and spectrum, acute in vivo toxicity, absorption, and so on, can be used to select new drug candidates for further development. The in vitro assay provides a rapid, sensitive means to simultaneously compare several compounds, using only milligram quantities of each drug. Such an in vitro system can predict only the inherent toxicity of the drug to hepatocytes. It cannot predict the modifying influences of the rate of absorption, distribution, and elimination that obtain in vivo. Hepatocytes used in such assays do have the ability to metabolize drugs to a limited extent. However, once the drug is placed into the culture medium, the cells are essentially exposed to that concentration of drug for the duration of the incubation period. Such prolonged exposure at the same concentration may not occur in vivo; therefore, the toxicity observed in vitro may not be predictive of the in vivo situation. The same arguments hold for other cell systems such as renal tubule cell, myocardial cell, or keratinocyte cell cultures. However, cell types such as the latter two have the additional limitation that they do not metabolize drugs as do liver or renal tubule cells. If the metabolite of a drug causes toxicity in these cell types, this would be missed in an in vitro assay unless some provision is made for introducing drug metabolizing enzyme systems into the assay.

Prescreening new compounds in in vitro tests for target organ toxicity can minimize the severity of toxicity and thus suffering and mortality in subsequent animal toxicity studies. Because fewer compounds will be tested in animals with this approach, in vitro tests will reduce the numbers of animals used in drug development.

To use such tests rationally, one needs to know the expected target organ for the type of drug under study. This knowledge may come from experience with similar drugs or from initial pharmacological or toxicological studies with a series of new drugs. Cell types such as the rat hepatocyte can also be used to study the mechanism of drug-induced hepatotoxicity. For instance, in addition to measuring the loss of intracellular enzymes, one could measure the amounts of intracellular glutathione, adenosine triphosphate (ATP), calcium, or enzymes of the various metabolic pathways to gain insight into possible mechanisms of toxicity.

Primary cultures of hepatocyte and tubule cells are prepared by collagenase perfusion and dissociation of the liver or kidney into cell suspensions. Primary rat myocardial cells can be similarly prepared by using trypsin to dissociate the cells. Rhythmic contractions of myocardial cells can be observed in such cultures. Changes in these contractions serve as another end point for the measurement of toxicity (27). The L6 rat skeletal muscle myoblast cell line has been used to screen compounds for potential muscle irritating and damaging activity (28).

Newer technologies that use the methods of molecular biology are being applied to rapid screening of large numbers of compounds (high-throughput screening) for potential toxicity. These methods make use of knowledge of selected genes or patterns of gene expression that have been associated previously with toxic responses. The methods are known generally as genomics or proteomics. One approach is to expose cells, tissues, or an animal to a test compound and extract either RNA or proteins from the cells or tissue of interest. If RNA is extracted, DNA complementary to this RNA (cDNA) can be applied to a chip containing DNA from thousands of different genes to determine which of those genes are expressed in the drug-exposed tissues. This is determined by labeling the cDNA as it is formed with either radioactivity or a fluorescent molecule. The site on the chip where radioactivity or fluorescence is detected allows determination of genes that had expression levels either higher or lower than in untreated tissues. If protein is extracted, the protein extract is subjected to multidimensional gel chromatography to determine patterns of alteration in gene expression. These patterns are then compared to patterns previously shown to be associated with toxic responses in vivo with other compounds. Efforts are being made in some cases in the pharmaceutical industry to understand toxic responses in humans better by determining genetic susceptibility to adverse effects of specific drugs. Gene expression or genetic polymorphisms are studied in people who have had adverse reactions to the drug of interest. Once the genetic nature of the drug sensitivity is understood, other potential patients can be tested to avoid those who might be sensitive to adverse effects of the drug.

There are a few in vitro models available for rapid testing of ocular irritation potential. These tests were developed to mimimize exposure of live animals to potentially distressing procedures. These include the BCOP (Bovine Corneal Opacity) test and the Eyetex test (29, 30). In the BCOP test, bovine corneas obtained from slaughter houses are treated under tissue culture conditions with solutions of the test compound to determine whether corneal opacity occurs. Corneal opacity is considered a type of severe ocular irritation. In the Eyetex test, a the test compound is added to a solution of proprietary proteins. The protein solution is then observed for cloudiness or precipitation. Both these tests have a fair record of concordance with in vivo data.

Tests such as these are useful for rapid screening of large numbers of compounds or for preliminary testing prior to testing in animals. Compounds predicted to cause severe irritation presumably would not be tested further in animals.

In vitro systems have also been developed to use as screening tools or for mechanistic studies of teratogenicity. Among the most commonly cited examples are the limb bud assay (31) and the hydra assay (32). In the limb bud assay, cultures of mouse embryonic limb bud mesenchymal cells are prepared and exposed to test compound. Under normal conditions, the mesenchymal cells differentiate into chondrocytes. An agent that disrupts this differentiation process would be expected to be teratogenic. In the hydra assay, two protocols have been recommended. In both cases, the ratio of the test compound's degree of toxicity for the adult form of the organism to that for an undifferentiated form of the organism is the end point used to determine the relative teratogenic potential. In one protocol, the undifferentiated form is a mass culture of dissociated cells. In the other, it is the digestive region dissociated from the adult form. In both cases, the undifferentiated form will differentiate or regenerate the adult form unless inhibited by the test compound. Neither of these systems have found wide usage in new drug development. The hydra assay is probably better suited to screening environmental contaminants because of the economy of the assay and the massive number of environmental chemicals likely to be screened.

Immunotoxicity

Immunotoxicology is an area where refined methods of measuring functional changes in a particular organ system are increasingly being applied to the safety assessment of new compounds. The assays used for functional measurements are taken from methods used for research in immunology and for screening new compounds for in vivo antimicrobial or antitumor efficacy.

In routine subchronic-chronic toxicity studies, several variables are obtained that can be indicators of immunotoxicity. These include differential white blood cell count, spleen and thymus weights, histological assessment of spleen cellularity, and thymus and lymph node morphology. However, the new, more refined methods being applied to immunotoxicity assessment are more specific; they can indicate whether particular cell types or immunological functions are affected. These newer specific assays are not at present being applied on a routine basis to screen new drugs. Instead, they are applied selectively in cases where the particular drug is suspected of being an immunotoxicant. This suspicion may arise from a structural relationship to previously known immunotoxicants (e.g., antivirals or steroids) or from data obtained in routine toxicity studies.

The National Institute of Environmental Health Sciences National Toxicology Program has recommended that immunotoxicity testing be done using a tier approach (33). Inbred strains of mice are usually used for these assays although methodology has been developed for conducting most of the assays in rats (34). The first tier would include those indicators routinely obtained in subchronic/chronic toxicity studies mentioned above, as well as specific assays for humoral-mediated, cell-mediated, and nonspecific immunity. In these specific assays, the animals are treated with a test compound for 14 days, for example, then their peripheral blood lymphocytes or spleen cells are used in the assay. The length of time for in vivo drug treatment will vary with the type of drug and its pharmacokinetics. The assay for humoral-mediated immunity consists of the plaque-forming cell assay to determine the numbers of splenic cells forming IgM antibody to a T-Cell dependent antigen (e.g., sheep red blood cells). In this assay, spleen cells from the treated animal are incubated in a mixture of complement and sheep red blood cells. The number of antibody-forming cells is determined from the plaques formed by areas of lysed sheep red blood cells on a background of unlysed red blood cells. The recommended assays for cell-mediated immunity are lymphocyte blastogenesis in response to a mitogenic stimulus (concanavalin A or phytohemagglutinin) and the mixed leukocyte response. In these assays, the extent of lymphocyte cell division is determined by incorporation of radiolabeled thymidine into cellular DNA after stimulation by a mitogen or by allogeneic lymphocytes, respectively. Natural killer cell activity is determined by incubating spleen cell suspensions from drug-treated animals with radioactive chromium-labeled target cells such as YAC-1 tumor cells. Release of free chromium into the culture medium is a measure of cell-killing activity.

The second tier includes assays that help determine the specific cell type involved in a response seen in the first tier and assays designed to assess the in vivo relevance of results seen in the first tier. These assays include quantitation of the numbers of splenic B and T lymphocytes, IgG humoral-mediated immunity, cell-mediated and nonspecific immunity, and host-resistance challenge models. Quantitation of B and T cells is done using cell-specific antibody techniques. IgG humoral-mediated immunity is assessed similarly to IgM humoral-mediated immunity (discussed above) except that a longer in vitro incubation time is used to allow the development

of IgG antibodies. Cell-mediated immunity can be assessed by the interleukin 2-dependent expression of cell-killer activity in a manner similar to the natural killer cell assay mentioned above. Nonspecific immunity can be assessed by counting the numbers of peritoneal macrophages in drug-treated animals and by measuring the ability of these macrophages to phagocytize radioactive chromium-labeled chicken red blood cells.

The host resistance models provide a means to directly assess the functional, and clinical significance of any changes found in the preceding tests. Direct assessment is important because of the multifactorial nature of immune responses and because of the well-known reserve capacity of the immune system. These assays consist of infectious agents or tumor cells that will routinely kill 10 to 30% of the animals not treated with drugs. An increase in the lethality of the challenge agent indicates suppression of the immune response. The challenge agent can be selected so as to challenge either humoral or cell-mediated immune systems.

SAFETY EVALUATION OF BIOTECHNOLOGY-DERIVED PRODUCTS

The safety evaluation of biotechnology-derived products is a developing area, which has been the subject of controversy. Safety evaluation has been developed for each product on a case-by-case basis. The ICH and FDA have developed guidelines for the safety evaluation of proteins and peptides produced via biological or cell culture systems (35). Several issues that should be addressed for each product have been identified (36, 37), through discussion among toxicologists concerning the appropriateness of applying the toxicity protocols used for synthetic pharmaceuticals to peptide products of biotechnology, and from examples of the few biotechnology-derived products that have been developed for clinical use.

Much discussion has arisen from the issue of potential contamination of the final dosage form by cellular material from the cell culture system used to produce the product. If a mammalian system has been modified by insertion of a DNA vector into cellular DNA, the concern would be over the potential contamination of the final product by transforming DNA. This might potentially increase the risk for developing cancer in a patient receiving such a contaminated product. For such products, assays for transforming activity of the final product or of DNA extracted from the cell culture system would be appropriate. The target limit for contaminant DNA that has generally been discussed is 10 picograms of DNA per final dose. Assay for contaminant DNA should be established as a routine quality control procedure for production batches. Tests for transforming activity (e.g., cell transformation in C3H/10T 1/2 or Balb/c 3T3 cells) should be performed initially once the cell culture system is established and at periodic intervals to assure genomic stability of the cell culture system. Another source of problems with potential contamination is the antigenicity of the final product due to foreign cellular materials from the biological system used to produce the product, whether bacterial or mammalian. Also, with bacterial production, systems contamination with pyrogenic cell components must be avoided.

The development of neutralizing antibodies against a human protein that would be foreign to the test species is a major concern. The presence or absence of a neutralizing antibody must be determined in each test species used. This could very well limit the species that are appropriate to use in safety studies, and it could limit the duration of any toxicity studies conducted. Generally, neutralizing antibody titers could develop as early as 10 to 14 days after initiation of treatment. Therefore, for a product inducing such antibodies, toxicity studies of more than approximately 14- to 16-days duration would not be appropriate.

Knowledge of the product's pharmacological activity in the species used for safety testing is requisite for a valid toxicity study; obviously, this should be determined beforehand. The only toxicity induced by the product may be in the form of exaggerated pharmacological effects. Exaggerated pharmacological activity may arise through systemic exposure to a protein normally present in only small quantities at specific sites and/or through stimulation of nontarget receptors. In the latter case, if an inappropriate species is used, the safety of the product may be overestimated. Monoclonal antibody products represent a special case of this requirement. The species selected for testing a monoclonal antibody product should provide antigenic receptors for the antibody on the same target tissues as in humans. If the species selected has antigenic receptors that bind the antibody to unintended tissues, or if there are no antigenic receptors that bind the antibody, the results obtained from that species will be misleading and inappropriate. Generally, the closer in structure the product is to the natural product, the lower the need for extensive long-term toxicity studies; however, the greater the difference in structure between the biotechnology-derived product and the natural product, the more the situation comes to resemble that of a synthetic drug where more extensive toxicity studies are required. Improved predictions of human safety may be obtained through the use of transgenic animals bearing the human receptor. Alternatively, toxicity studies in animals may be

conducted using the animal homologue of the human protein under consideration. In addition, it may be possible to conduct studies in an animal model of the human disease. In these cases it should be demonstrated that the pharmacological response of the test system would be similar to the response in human tissues.

Much information can be gained from in-life physiological measurements such as blood pressure, electrocardiograms, body temperature, and respiratory rate. When the number of studies is reduced due to the lack of appropriate species and the potential for development of neutralizing antibodies, as much information as possible should be obtained from the few studies that are conducted. In-life measurements assume more importance in these cases. When the product is likely to contain proteins that may cause acute "serum sickness" type responses and pyrogens, measurements of blood pressure, body temperature, and respiratory parameters are especially important.

Local injection site irritation studies would be expected with these types of products because they are almost always given by injection.

Mutagenicity studies may be conducted if mutagenic contaminants are suspected. If a biotechnology-derived product is closely related to the natural product, there would be no need to conduct mutagenicity studies to assess mutagenicity of the compound per se. Carcinogenicity studies are not routinely needed for these types of products unless there is a concern about induction of tissue proliferation (e.g., with growth factors) or immunosuppression. Reproductive toxicity studies may or may not be conducted, depending on the nature of the product and its intended use. If the compound is suspected of having abortifacient actions (as with the interferons), a study in pregnant primates would be indicated.

USE OF METABOLISM AND PHARMACOKINETIC DATA IN DRUG SAFETY EVALUATION

There are a host of factors that can affect blood and tissue levels of a drug and thus affect exposure of target organs (38, 39).

Selection of the route of administration can affect the degree of absorption and time course of drug levels. The intravenous route can give a rapid rise and fall in drug levels following bolus adminstration or a steady level during slow infusion. Oral administration usually gives a slower rise and fall in blood levels than intravenous dosing does; however, the extent of absorption may be the same, that is, 100%. Intramuscular, intraperitoneal, and subcutaneous administrations generally give rates of rise and fall of blood levels between those of intravenous and oral administration. If a drug is insoluble, or precipitates out of solution after injection, the intramuscular, intraperitoneal, and subcutaneous routes may give rise to a prolonged period of absorption from the injection site until all the drug has become solubilized and absorbed from the site.

For the oral route of administration, various factors such as the use of solutions versus suspensions versus dietary administration, the selection of the vehicle, the particle size, and the particle dissolution rate can affect the rate and extent of absorption. The absorption of a drug is likely to be greater if a solution rather than a suspension can be administered; on the other hand, the time course of rise and fall of drug blood levels is likely to be more prolonged with a suspension. There is some evidence that increasing the concentration of the suspending agent, for example, methylcellulose, can cause slower and less complete absorption. With dietary administration, the rate and extent of absorption are generally expected to be lower than with gavage dosing. Because rodents feed nocturnally, the period of drug consumption will be markedly prolonged compared with gavage dosing. However, because of the degree of dispersion necessary to prepare homogeneous dietary drug mixtures, and because of the long period available for drug administration, absorption from the diet can often approach 100% of that for gavage dosing. If a drug is intended to be administered several times per day, for example, over a 12-h period, dietary administration may provide the most appropriate means of modeling the expected human exposure. Particle size and the rate of particle dissolution can affect the rate and extent of drug absorption. The smaller the particle size, the greater the surface area per mass of the particle, and therefore the greater the rate of dissolution and absorption. The toxicologist has to be careful about using different batches of the same compound because the particle size may differ among batches. Such variation may affect the blood levels of a drug obtained in the test animals, which may in turn affect the toxicity observed.

For many drugs, most absorption probably occurs in the small intestine due to the vast surface area available. The rate of gastric emptying affects the delivery of drug to the small intestine. Therefore, the gastric emptying rate may affect the rate of absorption. The normal gastric emptying times are about 10 min for rodents, 30 min for rabbits, and 1.3 h for dogs. However, these times may be altered by drug administration, either pharmacologically or by the effect of variables such as dose volume or osmotic pressure. This further complicates the effect of gastric emptying on the absorption kinetics of the drug.

Oral absorption may be decreased as a percentage of administered dose at high doses. This can occur if the gastrointestinal transit time is less than the time required for complete absorption. This is sometimes the case with poorly absorbed drugs and drugs with poor solubility.

The selection of the vehicle can make a dramatic difference in the toxicity of a compound. An example would be the use of corn oil instead of methylcellulose to suspend a lipophilic compound. Corn oil may allow a better suspension, and possibly dissolve more of the drug, but it may also delay absorption of the drug due to retention in the lumen of the gastrointestinal tract. Opposite effects can also be obtained. I experienced a case where the toxicity of a drug was markedly greater when suspended in corn oil than when suspended in methylcellulose. In attempting to conduct intravenous studies with drugs having poor aqueous solubility, it is often necessary to use nonaqueous solvents, for example, ethanol, propylene glycol, or polyethylene glycol, to dissolve the drug. All nonaqueous solvents cause some degree of toxicity, whether cardiotoxicity, hemolysis, hepatic toxicity, or renal toxicity, that may complicate the toxicity assessment of the drug. Therefore, nonaqueous vehicles must be carefully selected, and the amounts and concentrations administered must be kept to a minimum.

The rate of intravenous injection can be a critical factor in determining the observed toxicity. Rapid administration of intravenous formulations will give high peak blood levels of drug, which may cause toxicity that could be avoided by infusing the drug at a slower rate. It is important in comparing the results of intravenous toxicity studies to note the injection rate. In designing a series of studies with a given formulation, it is important to use the same rate of injection. When determining upper tolerated limits for volume and rate of injection for a given species, knowledge of the glomerular filtration rate for the species can provide a useful guide. If one knows the rate of injection that will be used clinically, it is prudent to use the same injection rate in the animal toxicity studies. Interspecies differences in the rate and extent of absorption and in the metabolism of drugs are common and have a major effect on observed toxicity. Unfortunately, the toxicologist usually does not have this information when designing the initial toxicity studies. If the rat or dog is found to be an inappropriate species due to major differences in these parameters when compared to humans, some of the toxicity work may have to be repeated in another species. With some drugs, a large percentage may be metabolized to an inactive form by the intestinal mucosa or the liver before all the absorbed drug reaches the systemic circulation. This is called the "first-pass effect" because some of the active drug is removed during the first pass through the mucosa or liver, which obviously decreases the effective or apparent absorption of the drug. On the other hand, if a large amount of the drug is given (e.g., at high doses in a toxicity study) so that the capacity of the first-pass metabolism is exceeded, the blood levels obtained will be disproportionately high and lead to excessive toxicity. It is important to be aware of any first-pass metabolism when designing oral or intraperitoneal toxicity studies. When a drug is administered intraperitoneally, most of the drug is absorbed via the portal circulation that flows through the liver before reaching general circulation.

When an active drug is excreted in the bile, it may be reabsorbed. This sets up a cycle of biliary excretion, reabsorption, biliary excretion, and reabsorption. Usually only a fraction of the drug is excreted in the bile, and only a fraction thus excreted is reabsorbed. Therefore, the cycle does not continue indefinitely when treatment is stopped. However, such enterohepatic recirculation can lead to higher blood levels of drug than would otherwise be obtained and can prolong the apparent blood or tissue half-life of the drug. Such a cycle can lead to alterations in the blood level-versus-time curve such that after a single dose the initial peak level is followed by a smaller peak caused by the reabsorption of excreted drug from the small intestine. This phenomenon can affect the blood levels and thus the toxicity obtained with a given dosage level, and should be taken into consideration when designing toxicity studies.

With substantial increases in body weight and in pregnancy, the volume of distribution will be increased since the blood and tissue volumes increase. The half-life ($t_{\frac{1}{2}}$) is directly related to the volume of distribution (V_d) as follows:

$$t_{1/2} = \frac{V_d \times AUC \times 0.693}{D}$$

where AUC is the area under the blood concentration-versus-time curve and D is the dose. With substantial increases in body weight, the $t_{\frac{1}{2}}$ for elimination increases also. The time required to reach a steady-state concentration of drug in blood or tissues upon repeated administration, or to eliminate virtually all of a drug after cessation of treatment, is roughly 4 to 5 $\times$ $t_{\frac{1}{2}}$. Therefore, the weight gain found in rodents during their phase of rapid growth (the first 3 to 5 months and during pregnancy) will have an effect on the time required to reach steady-state blood–tissue levels, and on the time required to clear a drug from the body.

The concept that four to five half-lives are required to achieve a steady blood level is also of practical importance when determining whether there is any accumulation of

blood levels over the course of a subchronic-chronic toxicity study. It is the usual practice to take blood samples early in the study and again toward the end of the study to determine if blood levels of the drug are higher at the end than at the beginning of treatment. If not enough time is allowed at the beginning of the study for steady-state blood levels to be obtained, the results may mislead one to conclude that accumulation has occurred. Similarly, if a recovery period is included in the study design, it is important to allow enough time for the drug to be eliminated from the test animals and to allow additional time for recovery from drug effects beyond that required to eliminate the drug.

Changes in the organs of elimination (e.g., liver and kidneys) can have dramatic effects on the blood levels of a drug, such that excessive blood levels and concomitant toxicity are obtained due to reduced elimination of drug. However, some drugs induce increased synthesis of metabolizing enzymes, mainly in the liver, and sometimes increased rate of bile flow. In these cases, lower-than-expected blood levels may be obtained. This may occur at higher dosages and after a period of time necessary to stimulate the metabolic machinery.

Many drugs are metabolized to products that have pharmacological and/or toxicological activity. The pharmacokinetics of these metabolites may differ from those of the parent drug, complicating the design and interpretation of toxicity studies.

EXTRAPOLATION OF ANIMAL TOXICITY DATA TO HUMANS

Continually present in drug safety evaluation are questions concerning the predictive value of toxicological findings in animals relative to those expected to occur in humans. Two questions are of importance: Will the same kinds of toxic changes occur in humans that occur in laboratory animals? At what dosage may toxic changes be expected in humans relative to the dosage in animals?

There are differences in the processes of absorption, metabolism, and excretion that have been identified between several laboratory species and humans (40). These differences are useful in the selection of the appropriate species in which to study particular compounds. However, having selected, as far as practical, the appropriate species, it is generally recognized that if toxic changes occur in more than one species, they can be expected to occur in humans as well. Predicting the kind of toxicity that will occur in humans is usually less of a problem than predicting the dosage at which it will occur.

Generally, laboratory animals are less sensitive on a per body weight basis to the toxic effects of chemicals than are human beings. The sensitivity generally varies inversely with the size of the animals. More specifically, the sensitivity varies inversely with the surface area-to-body weight ratio. The surface area of a mammal is roughly the body weight raised to the two-third power. The ratio of the body surface area to the body weight determines the amount of energy that must be expended to maintain normal body temperature. The energy that is converted to body heat to maintain temperature is derived from the basal metabolic processes of the animal. The smaller the animal, the higher the metabolic rate because the ratio of surface area to body weight is larger. The rate at which drugs and other chemicals are metabolized and excreted generally varies with the basal metabolic rate of the species, which in turn varies with the surface area-to-body weight ratio. Thus, dosages comparable to a 1-mg/kg dosage in humans, after adjustment for differences in surface area, would be about 3 to 5 mg/kg for species such as rabbits, dogs, and monkeys, and about 10 mg/kg for rodents such as rats and mice. These factors are frequently used to establish the safe dosage for humans on the basis of animal toxicity studies. For example, the no-toxic-effect dosage obtained in a rodent toxicity study may be divided by 10 to determine the maximum dosage for clinical use in humans.

Application of the results of carcinogenicity studies to humans continues to be a controversial subject. Some of the issues involved were mentioned in the section on carcinogenicity bioassays. The relevance of tumors found at unrealistically high dosages is questioned as is the extrapolation of tumor incidence–dose response curves to low dosages where, presumably, no additional tumors above spontaneous incidence would be found. In some cases, as mentioned previously, where there is a clear change in metabolism or pharmacokinctics of the drug at high dosages, results found at such dosages may not be expected to occur in humans under normal, presumably much lower, exposure levels. The validity of the extrapolation of tumor incidence-dose response curves to low dosages where, theoretically, the tumor incidence should be zero, depends upon whether the compound acts as a genotoxic initiator or as a nongenotoxic promotor or cocarcinogen. Genotoxic compounds induce initiating, that is, DNA or chromosome damaging, events that can eventually lead to tumor formation. Nongenotoxic promoters or cocarcinogens can lead to tumor formation only in conjunction with the action of a genotoxic initiating compound. Promoters can act when applied

either at the same time as or after the application of the initiating compound. Cocarcinogens can act only when applied at the same time as the initiator. For most genotoxic compounds, theoretically, one molecule is sufficient, if it reaches the target DNA, to cause an initiating lesion. For promoters and cocarcinogens, it is generally understood that one molecule is not sufficient; there is a threshold concentration or dose below which the promoting or cocarcinogenic action is not obtained. Therefore, extrapolation to dosages below the threshold dosage would be appropriate for agents that act via promotion or cocarcinogenesis but not for genotoxic initiating compounds. An example of a class of compounds that cause DNA damage indirectly via interaction with a non-DNA target would be the topoisomerase inhibitors. Through inhibition of topoisoermase action DNA strand breakage normally caused by the enzyme is maintained, whereas without the action of the drug the breaks would be rejoined as part of the normal process of maintaining the DNA. The interaction of drug with the topoisomerase enzyme is subject to a dynamic equilibrium so that a certain concentration of drug is required to interfere with enzyme action. This drug effect, although causing DNA damage should demonstrate a threshold effect; since concentrations of drug below the critical level should be without biological effect (41).

Application of animal data on eye, skin, muscle, and vein irritation and pain on injection to humans merits comment. The rabbit eye is the most commonly used model for predicting eye irritation. This model is generally believed to be slightly more sensitive than the human eye to most irritants (42, 43). The same can be said for the rabbit skin as a model for human skin irritants. The muscle irritation test in rabbits seems to have a good correlation with clinical experience regarding painful intramuscular formulations even though this is not a test for pain (44). If a formulation causes muscle damage in rabbits, because of the similarity of muscle tissue across species, it can be expected to do so in humans. Vein irritation studies in animals are not always predictive of clinical results. If a formulation produces marked vein irritation in animals, it will likely do so in humans. However, a low level or absence of vein irritation in animals does not mean the formulation will not be irritating clinically. Several variables influence the irritating nature of an intravenous formulation. These are concentration of the drug, tonicity, pH, formulation ingredients other than the drug, rate of infusion, and number of repeated injections in the same site. Naturally, any changes in these variables between animal and clinical studies affect the predictability of animal vein irritation studies. Results of animal muscle and vein irritation studies cannot be used to predict pain on injection in humans because pain is not an end point in these tests. The only available models for pain on injection, by any route, are the rat paw lick and mouse scratch tests mentioned previously. These seem to be predictive for the few compounds that have been tested for a correlation with clinical experience.

The application of data from reproductive toxicity studies to humans involves special considerations in addition to those mentioned above for general toxicity studies. This is especially true of teratology studies. The actual reliability of animal models to predict toxicity in this area is never known until adequate epidemiological studies are conducted in humans. Such epidemiological studies are rarely conducted due to the expense and time required. The ideal animal model would handle the drug (i.e., through absorption, distribution, metabolism, excretion, and maternal fetal distribution) identically to humans and would have the same maternal–placental–fetal relationships as humans. Because the animal models available to us rarely represent the ideal model, the best compromise is to test, for teratogenicity at least, in two or more species (45). If the drug causes fetal toxicity or teratogenicity in two or more species, it would be a good assumption that the drug will do the same in humans. Of the chemicals known to be teratogenic in humans, all but one class, the coumarin anticoagulants, are also teratogenic in at least one species of laboratory animal, (46). However, there are many chemicals known to be teratogenic in animals that are apparently not teratogenic in humans. This may be due to lower levels of exposure in humans or to more careful management of exposure as in the case of pharmaceuticals. Animal models have greater predictive power when combinations of two or more species are used. Among the species commonly used for teratogenicity testing, rats and mice seem to be the best for modeling human responses. However, the use of rabbits in combination with either rats or mice provides greater predictability because rabbits are less likely to yield false positive results. The rabbit was originally selected as a second, nonrodent species for routine teratogenicity testing because of its responsiveness to thalidomide.

Finally, it should be pointed out that although the animal models commonly used can predict the fetal toxicity of a drug in humans, the exact type of fetal toxicity in humans cannot be predicted. If an agent causes fetal toxicity of any kind in animals, it should be suspected as a potential human teratogen or fetal toxin. Valproic acid serves to illustrate this concept: It causes skeletal abnormalities in rats and rabbits. In humans, it has been shown by epidemiological studies to cause a low incidence

of spina bifida, a neural tube defect. Therefore, the type of fetal toxicity caused in the animal models may different from that caused in humans, but the fetal toxicity in humans could be predicted from the animal studies. It turns out that valproic acid also causes neural tube defects in mice and hamsters. In summary, the animal models used for teratogenicity screening seem to provide reasonable predictability for fetal toxicity in humans. The greater the number of species with positive results, the greater the likelihood that the drug will also cause fetal toxicity in humans.

REFERENCES

1. Sherman, M. Drug and Cosmetic Industry. **1988**, *142* (2), 37–78.
2. McCutcheon, R.S. Toxicology and the Law. *Casarett and Doull's Toxicology, The Basic Science of Poisons*, 2nd Ed.; J. Doull, J., Klaassen, C.D., Amdur, M.O., Eds.; MacMillan: New York, 1980; 727–733.
3. Finney, D.J. Probit Analysis. Cambridge University Press: Cambridge, MA, 1971.
4. Brownlee, K.A.; Dodges, J.L.; Rosenblatt, M. J. Am. Statist. Assoc. **1953**, *48*, 262–277.
5. Thompson, W.R. Bacteriol. Rev. **1947**, *11*, 115–145.
6. Celozzi, E.; Lotte, J.; Stapley, E.O.; Miller, A.K. J. Pharm. Meth. **1980**, *4*, 285–289.
7. Comereski, C.R.; Williams, P.D.; Bregman, C.L.; Hottendorf, G.H. Fundam. Appl. Toxicol. **1986**, *6*, 335–338.
8. Detweiler, D.K. The Use of Electrocardiography in Toxicological Studies with Rats. *The Rat Electrocardiogram in Pharmacology and Toxicology*; Budden, R., Detweiler, D.K., Zbinden, G., Eds.; Pergamon Press: Oxford, 1981; 83–115.
9. Schaeppi, U.; Liverani, F. Agents Actions **1977**, *7* (3), 347–351.
10. Fort, F.L.; Yamamoto, M.; Ando, T. The Toxicol**. 1987**, *7* (1), 96.
11. Vobig, M.A.; Klotz, T.; Staak, M.; Bartz-Scmidt, K.U.; Engelmann, U.; Walter, P. Retinal Side-Effects of Sildenafil. The Lancet **1999**, 353–375.
12. Tennant, R.W.; Margolin, B.H.; Shelby, M.D.; Seiger, E.; Haseman, J.K.; Spalding, J.; Caspary, W.; Resnick, M.; Staseiwicz, S.; Anderson, B.; Minor, R. Science **1987**, *236*, 933–941.
13. *International Conference on Harmonisation Draft Guideline on Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals*; Federal Register 1987; 16026–16029.
14. Galloway, S.M.; Deasy, D.A.; Bean, C.L.; Kraynak, A.R.; Armstrong, M.J.; Bradley, M.O. Mutat. Res. **1987**, *189*, 15–25.
15. Cifone, M.A.; Myhr, B.; Eiche, A.; Blocsfoldi, G. Mutat. Res. **1987**, *189*, 39–46.
16. Haseman, J.K. Fundam. Appl. Toxicol **1985**, *5*, 66–78.
17. Clayson, D.B. Mutat. Res. **1989**, *221*, 53–67.
18. Peto, R.; Br., J. Cancer . **1974**, *29*, 101–105.
19. Peto, R.; Pike, M.C.; Day, N.E.; Gray, R.G.; Lee, P.N.; Parish, S.; Peto, J.; Richards, S.; Wahrendorf, J. *IARC Monograph on Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal, Suppl. 2*; IARC: Lyon, France, 1980.
20. McKnight, B. Fundam. Appl. Toxicol. **1988**, *10*, 355–364.
21. Arakawa, M.; Hori, T.; Innoue, S.; Kanamura, K. Jap. Pharmacol. Therap. **1985**, *13* (7), 459–469.
22. Campbell, D.H.; Garvey, J.S.; Cremer, N.E.; Sussdoff, D.H. Gel Diffusion. *Methods in Immunology*, 2nd Ed.; W.A. Benjamin, Inc.: New York, 1970; 250–260.
23. Harber, L.C. J. Invest. Dermatol. **1981**, *77*, 65–70.
24. Sams, W.M. Arch. Dermatol. **1966**, *94*, 773–777.
25. Fort, F.L.; Kotz, R.V. The Toxicol. **1988**, *8*(1), 129.
26. Ljunggren, B. Photodermatol **1984**, *1*, 96–100.
27. Miletich, D.J.; Albrecht, R.F.; Hoishouser, S.J. In Vitro Toxicol. **1987**, *1* (2), 77–84.
28. Williams, P.D.; Masters, B.G.; Evans, L.D.; Laska, D.A.; Hottendorf, G.H. Fundam. Appl. Toxicol. **1987**, *9*, 10–17.
29. Sina, J.F. Validation of the Bovine Corneal Opacity-Permeability Assay as a Predictor of Ocular Irritation Potential. In Vitro Toxicol. **1994**, *7*(3), 283–290.
30. Sina, J.F.; Galer, D.M.; Sussman, R.G.; Gautheron, P.D.; Sargent, E.V.; Leong, B.; Shah, P.V.; Curren, R.D.; Miller, K. A Collaborative Evaluation of Seven Alternatives to the Draize Eye Irritation Test Using Pharmaceutical Intermediates. Fundam. Appl. Toxicol. **1995**, *26*, 20–31.
31. Kistler, A. Arch. Toxicol. **1987**, *60*, 403–414.
32. Wilby, O.K.; Newall, R.D.; Tesh, J.M. Food Chem. Toxicol. **1986**, *24* (6/7), 651–652.
33. Luster, M.I.; Munson, A.E.; Thomas, P.T.; Holsapple, M.P.; Fenters, J.O.; White, K.L.; Dauer, L.D.; Germolec, D.R.; Rosenthal, G.J.; Dean, J.D. Fundam. Appl. Toxicol. **1988**, *10*, 2–19.
34. Exon, J.H.; Koller, L.D.; Talcott, P.A.; O'Reilly, C.A.; Henninsen, G.M. Fundam. Appl. Toxicol. **1986**, *7*, 387–397.
35. International Conference on Harmonisation Guidance on Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals. Federal Register 1997; *62 222* 61515–61519
36. Graham, C.E. *Preclinical Safety of Biotechnology Products Intended for Human Use*; Alan R. Liss, Inc.: New York, 1987.
37. Fort, F.L. J. Biomed. Environm. Sci. **1988**, *1* (2), 223–224.
38. Smyth, R.D.; Hottendorf, G.H. Toxicol. Appl. Pharmacol. **1980**, *53*, 179–195.
39. Clark, B.; Smith, D.A. CRC Crit. Rev. Toxicol. **1984**, *12* (4), 343–385.
40. Calabrese, E.J. Drug Metab. Rev. **1984**, *15* (3), 505–523.
41. Fort, F.L. Mutagenicity of Quinolone Antibacterials. Drug Safety **1992**, *7* (3), 214–222.
42. Freeberg, F.E.; Nixon, G.A.; Reer, P.J.; Weaver, J.E.; Bruce, R.D.; Griffith, J.F.; Sanders, L.W. Fundam. Appl. Toxicol. **1986**, *7*, 626–634.
43. Kay, J.H.; Calandra, J.C. J. Soc. Cosmetic Chem. **1962**, *13*, 281–289.
44. Gray, J.E. Fundam. Appl. Toxicol. **1981**, *1*, 290–292.
45. Schardein, J.L.; Schwetz, B.A.; Kenel, M.F. Environ. Health Perspect. **1985**, *61*, 55–67.
46. Schardein, J.L.; Keller, K.A. CRC Crit. Rev. Toxicol. **1989**, *19* (3), 251–339.

BIBLIOGRAPHY

Acute Eye Irritation Testing, Recommended Guideline; Interagency Regulatory Liaison Group: Testing Standards and Guidelines Work Group, January 1981.

Acute Dermal Toxicity Test, Recommended Guideline; Interagency Regulatory Liaison Group: Testing Standards and Guidelines Work Group, January 1981.

Acute Oral Toxicity Testing in Rodents, Recommended Guideline; Interagency Regulatory Liaison Group: Testing Standards and Guidelines Work Group: January 1981.

Alder, S.; Zbinden, G. *National and International Drug Safety Guidelines*; M. T. C. Verlag: Zollikon, Switzerland, 1988.

Brusick, D. *Principles of Genetic Toxicology*; Plenum: New York, 1980.

Brusick, D.J.; Bradley, M.O. In Vitro Toxicology **1988/89**, *2* (3), 151–236.

Graham, C.E. *Preclinical Safety of Biotechnology Products Intended for Human Use*; Alan R. Liss: New York, 1987.

Grice, H.C. *The Selection of Doses in Chronic Toxicity/Carcinogen Studies; Age-Associated (Geriatric) Pathology: Its Impact on Long-Term Toxicity Studies; Current Issues in Toxicology*; Springer-Verlag: New York, 1984.

Grice, H.C. *Interpretation and Extrapolation of Chemical and Biological Carcinogenicity Data to Establish Human Safety Standards; The Use of Short-Term Tests for Mutagenicity and Carcinogenicity in Chemical Hazard Evaluation; Current Issues in Toxicology*; Springer-Verlag: New York, 1984.

Hayes, A.W. *Principles and Methods of Toxicology*, 3rd Ed.; Raven Press: New York, 1994.

Heddle, J.A.; Dean, S.; Hohmi, T.; Boerrigter, M.; Casciano, D.; Douglas, G.R.; Glickman, B.W.; Gorelick, N.J.; Mirsalis, J.C.; Martus, H.J.; Skopek, T.R.; Thybaud, V.; Tindall, K.R.; Yajima, N. In Vivo Transgenic Mutation Assays. Environ. Mol. Mutagen. **2000**, *35*, 253–258.

International Conference on Harmonisation; Guideline on Dose Selection for Carcinogenicity Studies on Pharmaceuticals. Federal Register **1995**, 11278–11280.

International Conference on Harmonisation; Final Guideline on the Need for Long-Term Rodent Carcinogenicity Studies of Pharmaceuticals. Federal Register **1996**, *61, 42*8154–8156.

International Conference on Harmonisation; Guidance on Dose Selection for Carcinogenicity Studies on Pharmaceuticals: Addendum on a Limit Dose and Related Notes, Federal Register **1997**, 64259–64261.

International Conference on Harmonisation; Guidance on Testing for Carcinogenicity of Pharmaceuticals, Federal Register **1988**, *63 35,* 8983–8986.

Kesterson, J.W. Drug Inform. J. **1982, Jan/June**, *22–24.*

Kilbey, B.J.; Legator, M.; Nichols, W.; Ramel, C. *Handbook of Mutagenicity Test Procedures*; Elsevier: 1984.

Klaassen, C.D. *Casarett and Doull's Toxicology*, 5th Ed.; McGraw-Hill: New York, 1996.

MacGregor, J.T.; Farr, S.; Tucker, J.D.; Heddle, J.A.; Tice, R.R.; Turteltaub, K.W. New Molecular Endpoints and Methods for Routine Toxicity Testing. Fundamen. Appl. Toxicol. **1995**, *26*, 156–173.

Marzulli, F.N.; Maibach, H.I. *Dermatoxicology*, 5th Ed.; Taylor & Francis: Washington DC, 1983.

Marzulli, F.; Maguire, H.C. Food Chem. Toxicol. **1987**, *20*, 67–74.

Mirsalis, J.C.; Monforte, J.A.; Winegar, R.A. Transgenic Animal Models for Measuring Mutations In Vivo. Crit. Rev. Toxicol. **1994**, *24* (3), 255–280.

Naismith, R. Mutat. Res. **1987**, *189* (2), 81–183.

Nuwaysir, E.F.; Bittner, M.; Trent, J.; Barrett, J.; Afshari, C.A. Microarrays and Toxicology the Advent of Toxicogenomics. Mol. Carcinogen **1999**, *24*, 153–159.

Principles and Methods for Evaluating the Toxicity of Chemicals, Part I; Environmental Health Criteria No. 6, World Health Organization: Geneva, 1978.

Provost, G.S.; Kretz, P.L.; Hamner, R.T.; Matthews, C.D.; Rogers, B.J.; Lundberg, K.S.; Dycaico, M.J.; Short, J.M. Transgenic Systems for In Vivo Mutation Analysis. Mut. Res. **1993**, *288*, 133–149.

Purchase, I.F.H.; Hard, G.C. Toxicology In Vitro **1987**, *1* (1), 3–57.

Schardein, J.L. *Drugs as Teratogens*; CRC Press: Cleveland, 1976.

Schardein, J.L. *Chemically Induced Birth Defects*, 3rd Ed.; Marcel Dekker, Inc.: 2000.

Sontag, J.M.; Page, N.P.; Safflotti, U. *Guidelines for Carcinogen Bioassay in Small Rodents, National Cancer Institute Carcinogenesis, Technical Report Series No. 1*; Publication No. (NIH) 76-801: DHEW, February, 1976.

Taylor, P. *Practical Teratology*; Academic Press: New York, 1986.

Teratogenicity Studies in the Rat, Mouse, Hamster, or Rabbit: Recommended Guideline; Interagency Regulatory Liaison Group: Testing Standards and Guidelines Work Group, January 1981.

Wilson, J.G. *Environment and Birth Defects*; Academic Press: New York, 1973.